Chapter 5

Testing a Non-Iterative Genotype Probability Estimation Method

5.1 Introduction

Knowledge of true genotypes has the potential to improve genetic gain through more efficient selection of individual animals. Although not yet able to determine true genotype status of an animal for many traits of economic importance, genetic markers linked to those traits are beginning to be found. This marker information may be used to predict quantitative trait loci (QTL) genotypes (and therefore, genotypic value) of individual animals. This information will be useful in predicting genetic values and selecting parents to improve phenotypic performance of later generations. Parents will be able to be selected on their own genotype probabilities and mate allocation will be able to optimise expected progeny genotypes. For example, in the parental generation distinction could be made between homozygotes and heterozygote carriers of single recessive QTL alleles which are not distinguishable on phenotypic observations alone.

Currently, segregation analysis allows genotype probabilities at a single locus to be estimated using phenotype information on an individual and all its relatives and mates (van Arendonk *et al.*, 1989; Fernando *et al.*, 1993; Kinghorn *et al.*, 1993; Kinghorn and Kerr, 1995). Problems with segregation analysis arise when there are loops in pedigrees. A number of recursive and iterative methods have been developed to account for these loops (Janss *et al.*, 1995b; Kerr

and Kinghorn, 1995) or that involve cutting the loops (Stricker *et al.*, 1995). These methods only give approximations of the true likelihood. Janss et al (1995a) used Gibbs sampling to estimate fixed QTL effects and genotype probabilities of individual animals associated with those in a model that also included polygenic effects. Meuwissen and Goddard (1997) considered the same model, but used an iterative procedure, alternating between BLUP estimates of polygenic effects, and segregation analysis to determine genotype probabilities. Neither of these methods described uses marker information.

In principle, marker information could be included in all of the above methods to estimate genotype probabilities. However, all methods treat QTL effects linked to marker genotypes as fixed and such models require segregation analysis for genotype probabilities, implying complex analysis, and problems with loops and approximated likelihoods. Alternatively, QTL effects can be fitted as random in the estimation of breeding values. Information from linked marker genotypes is used to construct a gametic relationship matrix to represent a known variance-covariance structure among QTL effects (Fernando and Grossman, 1989). Estimated QTL effects from this model could be used to determine genotype probabilities. This would be a simple routine, which does not require iteration as in segregation analysis.

Kinghorn *et al.* (1993) used heights of normally distributed phenotypes to estimate probabilities of belonging to genotype classes. In this Chapter, this method is slightly adapted by using estimated QTL effects from the Fernando and Grossman (1989) model, rather than phenotypes to estimate genotype probabilities. Since pedigree information has already been included in the gametic relationship matrix, genotype probabilities can be calculated from an individual's estimated QTL effects directly, without considering family information. The methods can therefore also be seen as a simple extension of the procedure of Fernando and Grossman (1989) by transforming estimated QTL effects into genotype probabilities. Such a procedure is computationally much simpler than including marker information in the method of Meuwissen and Goddard (1997), since the latter needs iteration between BLUP and segregation type methods. The simple method may also be a simple alternative for segregation analysis for models with no polygenic effects and with no marker information.

The aim of the study was to propose and test a non-iterative method to calculate genotype probabilities. Simulation will be used to compare the method with methods using segregation analysis. Information on marker genotypes can be simply included in this method and we compare the value of marker information on accuracy of genotype probability estimation.

5.2 Method

5.2.1 Non-Iterative Method to Calculate Genotype Probabilities

The basis for the genotype probability calculations is a two-step procedure, as shown in Figure 5.1. The first step involves estimating random QTL effects (and possibly polygenic values) values using phenotypic observations as well as marker genotypes according to the model of Fernando and Grossman (1989). The second step involves calculating genotype probabilities for individual QTL conditional on the estimated QTL effects, and is an adaptation of the procedure proposed by Kinghorn *et al.* (1993). This method uses marker information in the first step to calculate probabilities of identity by descent of QTL alleles between gametes, with these probabilities included in the gametic relationship matrix to estimate breeding values. The methods assumes two alleles, and hance three genotypes, at the quantitative trait locus. Probabilities of individual animals belonging to each of the three classes are estimated in the second step. For each animal the expectation of its estimated QTL effect is calculated conditional on each possible QTL genotype. The mean and the variance of the estimated QTL effect per animal are then related to the genotypic means. The heights of this distribution at each QTL genotype are proportional to the genotype probabilities for this animal.

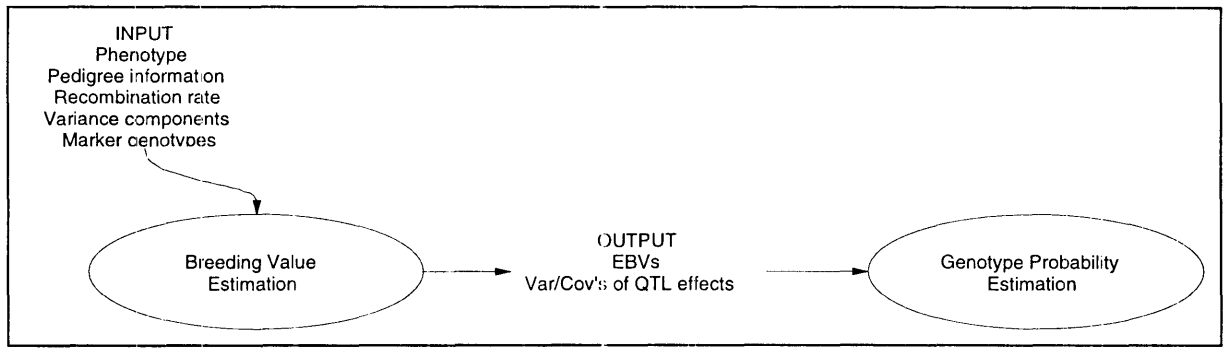

Figure 5.1 An illustration of the input and output parameters required for the 2-step genotype probability estimation method.

5.2.1.1 Estimation of Breeding Values Using Marker Information

Estimation of breeding values, which incorporate marker information, involves solving the following mixed model equations using the tabular method of Fernando and Grossman (1989) to build the inverse of the gametic relationship matrix. As given in Chapter Two, true breeding value (or genotype) of an individual (a_i) is composed of a polygenic effect (u_i) and two gametic effects (v_i) inherited from paternal (p) and maternal (m) alleles of its sire and dam.

The mixed inheritance model (incorporating both fixed and random effects) may be written in matrix notation as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \mathbf{e} \qquad (5.1)$$

where:

- $\mathbf{y}$ is the vector of phenotypic observations
- β is the vector of fixed effects
- $\mathbf{u}$ is the vector of random additive genetic effects due to loci not linked to the genetic markers
- $\mathbf{v}$ is the vector of random additive gametic effects at the marked QTL
- $\mathbf{e}$ is the vector of random residual effects

The matrices **X, Z** and **W** are known incidence matrices and the variance-covariance structure of random variables from Equation 5.1 is:

$$V\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}_u\sigma_u^2 & 0 & 0\\ 0 & \mathbf{G}_{v|r}\sigma_v^2 & 0\\ 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where:

- $\mathbf{A}_u$ is the numerator relationship matrix for the QTL unlinked to the marker loci
- $\mathbf{G}_{v|r}$ is the gametic relationship matrix for the marked QTL, given recombination rate(r)

I is an identity matrix

The total additive genetic variance maybe written as $\sigma_a^2 = \sigma_u^2 + 2\sigma_v^2$, where the QTL variance is $\sigma_q^2 = 2\sigma_v^2$. Letting $\alpha_u = \sigma_e^2/\sigma_u^2$ and $\alpha_v = \sigma_e^2/\sigma_v^2$, the following mixed model equations may be solved:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}_u^{-1}\alpha_u & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{G}_{v|r}^{-1}\alpha_v \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \tag{5.2}$$

Solving Equation 5.2 gives estimated breeding values for polygenic ($\hat{u}$) and QTL ($\hat{v}$) effects, as well as estimates of fixed effects ($\hat{\beta}$).

5.2.1.2 Genotype Probability Estimation

Estimation of genotype probabilities requires knowledge of the estimated QTL effects (as calculated above) and its standard error for each individual. The latter requires the inverse of the coefficient matrix in Equation 5.2. The gene frequency for the gene of major effect is assumed known in the base population (*p*). Probabilities of individual *i* having QTL genotype *q* were calculated as follows (from van Arendonk *et al.*, 1989):

$$\text{prob}(q_i) = \frac{\text{prior}(q_i)g(\hat{v}|q_i)}{\sum_{j=1}^{k} \text{prior}(v)g(\hat{v}|q_j)} \tag{5.3}$$

Where:

$prob(q_i)$	is the probability of individual *i* having genotype *q*,
$prior(q_i)$	is the prior probability of individual *i* having genotype *q* (simply based on gene frequency in the population)
$g(\hat{v}\|q_j)$	is the conditional probability of estimated QTL effect $\hat{v}_i = \hat{v}_i^p + \hat{v}_i^m$ given genotype q_j
$k = 3$	is the number of possible genotypes

Following the method of Kinghorn *et al.* (1993) $g(\hat{v}|q)$ was calculated as the height of the

normal distribution at estimated breeding value $\hat{v}$, with $\bar{v}_q$ equal to the mean of genotypes (q_i) means derived from known parameters a ,d and p and variance (σ^2) equal to the prediction error variance of the estimated QTL effects (from the inverted coefficient matrix). The BLUP estimates of QTL effect are regressed towards the mean, with the more accurate estimates regressed less than the less accurate estimates. The factor by which these estimates are regressed is the square of the accuracy of the estimated QTL effects (s^2). This requires estimates of QTL effects ($\hat{v}$) and the variance of QTL effects (σ^2) to be divided by (s^2) to de-regress the estimates to the expected genotype class means ($\bar{v}_q$).

$$g\left(\hat{v}\middle|q_j\right)=\frac{1}{\sigma\sqrt{2\pi}}exp\left(-0.5\frac{\left(\hat{v}/s^2-\bar{v}_q\right)^2}{\sigma^2/s^2}\right) \qquad (5.4)$$

This equation (5.4) gives heights at each of the three genotype class EBV distributions (Figure 5.2) which can then be compared relative to each other to give the probabilities of an individual belonging to the three genotype classes (Equation 5.3). The square of accuracy (s^2) and variance (σ^2) for use in Equation 4 are calculated using the variances and covariances between QTL effects which are the diagonal (variance) and off-diagonal (covariance) elements of the inverted coefficient matrix.

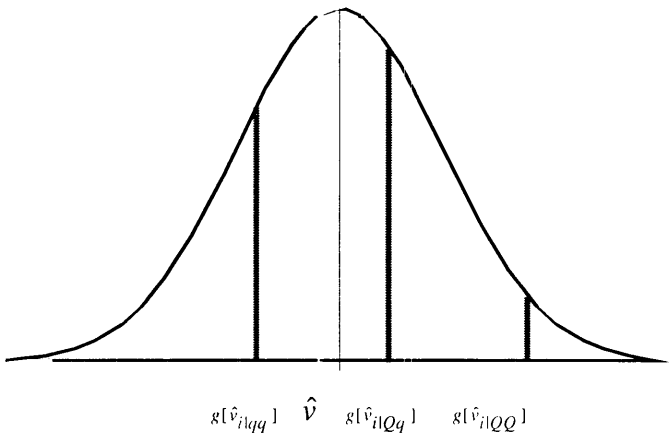

Figure 5.2 Heights of the distribution of $\hat{v}_i$ (EBV at QTL for animal i) at the expectations conditional on QTL genotype are proportional to genotype probabilities.

The genotype probability estimation procedure relies on use of the reliability of the estimated QTL effects. That is, the regression of true on estimated QTL effects multiplied by true QTL effects is used to calculate conditional probabilities (Equation 5.4). The coefficient matrix of the mixed model equations (Equation 5.2) maybe written as:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Where:

$$C_{11} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A_u^{-1}\alpha_u \end{bmatrix}$$
$$C_{21} = \begin{bmatrix} W'X & W'Z \end{bmatrix} \text{ and}$$
$$C_{22} = \begin{bmatrix} W'W + G_{v|r}^{-1}\alpha_v \end{bmatrix}$$

and a generalised inverse of the coefficient matrix is then:

$$\begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

The prediction error variance (PEV) of estimated QTL effects is the variance of the difference between the true and estimated QTL effects and is calculated as [$\text{var}(v - \hat{v})$]. It is the fraction of additive genetic variance not accounted for by the prediction. This was shown by Henderson (1975) to be:

$$PEV = C^{22}\sigma_e^2 = \left(1 - s^2\right)\sigma_a^2 \qquad (5.5)$$

Where: s is the squared correlation between the true and estimated QTL effects (accuracy). Equation 5.5 shows that the diagonal and off-diagonal elements of the inverted coefficient matrix for the animal equations are required to calculate the PEV for the animals. C_i^{22} contains, for each animal i, four elements referring to diagonal and off-diagonal elements of paternal and maternal QTL gametes.

For animal i the accuracy (s) can then be calculated, using the summed diagonal (and off-diagonal) elements of C_i^{22}, (where t has zeros and only a single one in positions of v_i^p and v_i^m in the vector **v**) as:

$$\begin{aligned}(1-r^2)\sigma_a^2 &= tC_i^{22}t'\sigma_e^2 \\ (1-r^2) &= tC_i^{22}t'\sigma_e^2/\sigma_a^2 \\ r &= \sqrt{1-tC_i^{22}t'\sigma_e^2/\sigma_a^2}\end{aligned}$$

And reliability can be calculated as the square of accuracy.

Using the same notation and using Equation 5.5 the standard error of prediction (σ_i) can be calculated as:

$$\sigma_i = \sqrt{var(v_i - \hat{v}_i)} = \sqrt{tC_i^{22}t'\sigma_e^2}$$

where the diagonal elements of the coefficient matrix are the variances and covariances between maternally (m) and paternally (p) inherited QTL effects (v_i):

$$\begin{aligned}\sigma_i^2 &= \text{var}(v_i^p + v_i^m) * \sigma_e^2 \\ &= \left(\text{var}(v_i^p) + \text{var}(v_i^m) + 2\,\text{cov}(v_i^p, v_i^m)\right) * \sigma_e^2\end{aligned}$$

5.2.2 Segregation Analysis Method of Fernando et al. (1993)

A non-iterative, recursive segregation analysis method was developed by Fernando *et al.* (1993). This method allows calculation of genotype probabilities for individual animals in a pedigree for a trait determined by genes at one or few loci. This method, however, cannot be used on pedigrees containing loops, because recursion is used. Recursion requires that probabilities of most distant relatives are calculated first and information from all relatives is used. This becomes an impossible task when loops are present in the pedigree. Janss *et al.* (1995b) developed 'iterative peeling', an equivalent to recursive peeling that can be used without modifications in looped pedigrees to obtain approximate likelihoods. Kerr and Kinghorn (1996) examined a similar iterative application of the Fernando *et al.* (1993) method which involves applying the genotype probability equations of Fernando *et al.* (1993) iteratively, rather than recursively, to calculate genotype probabilities for a pedigree containing loops. The adaptation by Kerr and Kinghorn (1996) is used in this study, however, the ordering of the pedigree as used by Kerr and Kinghorn (1996) was not implemented.

The mixed model fitted for analysis of combined segregation analysis for estimation of QTL probabilities and estimated breeding values for polygenic effects was the same as that used by Meuwissen and Goddard (1997). This model included QTL effects as fixed and involved weighting phenotypes by genotype probabilities. The general model for the analysis was:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{ZQq} + \mathbf{e} \qquad (5.6)$$

This is the same as Equation 5.1 except that random QTL effects **Wv** have been replaced by fixed QTL effects **ZQq**:

where:

- **q** is a (3*1) unknown vector of effects of the QTL genotypes (3 genotypes)
- **Q** is a (q*3) unknown incidence matrix with one at position (*j*,*k*) if animal *j* has genotype *k* and zeros elsewhere

The variance of polygenic effects is $Var(u) = G = A\sigma_u^2$ and $Var(e) = R = I\sigma_e^2$, where R is assumed to be diagonal and $Var(y) = ZGZ' + R$.

Equation 5.6, therefore, yields the following mixed model equations from the derivation of QTL, polygenic and fixed effects given in Meuwissen and Goddard (1997) and as explained in Chapter two of this thesis:

$$\begin{bmatrix} \mathbf{D} & \mathbf{W'X} & \mathbf{0} \\ \mathbf{X'W} & \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'W} & \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{q}} \\ \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{W'y} \\ \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} - \begin{bmatrix} r \\ 0 \\ 0 \end{bmatrix} \tag{5.7}$$

where:

$\lambda = \sigma_e^2 / \sigma_u^2$ and **W** is an (n*3) matrix of elements W_{ik}.

$$D = \begin{bmatrix} \sum_i W_{i1} & 0 & 0 \\ 0 & \sum_i W_{i2} & 0 \\ 0 & 0 & \sum_i W_{i3} \end{bmatrix}$$

and r is a (3*1) vector with elements $r_k = \sum_i W_{ik} \hat{u}_j(k)$ with j being the animal which produced record i.

Segregation analysis was used to solve for **W** following Meuwissen and Goddard (1997). This gave estimated genotype probabilities for each of the three genotypes.

5.2.2.1 Calculation of Genotype Probabilities

Following equations given by Fernando *et al.*(1993), the conditional probability of pedigree member i having genotype u_i, given all phenotypic data y is calculated as:

$$\Pr(u_i|y) = \frac{a_i(u_u)g(y_i|u_i)\prod_{j\in S_i} p_{ij}(u_i)}{L} \tag{5.8}$$

Where:

$a_i(u_i)$	is the 'anterior probability' for i having genotype u_i, this incorporates information from parents and full sibs
$p_{ij}(u_i)$	is the 'posterior probability' for i having genotype u_i , this uses information from progeny and is calculated as the product of terms through mates j of possible data set S_i
$g(y_i\|u_i)$	is the 'penetrance value', and uses information from self, it is the conditional probability of i having phenotype y_i given genotype u_i

The product of these probabilities gives the probability of genotype u_i for animal i, given all current and adjacent information. The sum of this joint probability over all genotypes u_i for i gives the likelihood for the pedigree *(L):*

$$L = \sum_{u_i} a_i(u_i) g(y_i|u_i) \prod_{j \in S_i} p_{ij}(u_i)$$

Equations for Calculation of Anterior and Posterior Probabilities

Using the following notation, the equations for the anterior and posterior probabilities required by Equation 5.8 can be calculated.

m and f	are parents of i
S_m and S_f	are the sets of all mates of m and f respectively
C_{mf}	is the set of all offspring of m and f and $s \in C_{mf}$
S_s	are the sets of all mates of each full sibling s
j	is the mate to i
S_j	is the set of mates of j
C_{ij}	is the set of offspring of i and j and $o \in C_{ij}$
S_o	are the sets of all mates of each offspring o

The equation to calculate an anterior probability $a_i(u_i)$ for i is:

$$
\begin{aligned}
a_i(u_i) = \quad & \sum_{u_m}\Bigg\{ a_m(u_m) g(y_m|u_m) \prod_{k\in S_m, k\neq f} p_{mk}(u_m) \\
& \times \sum_{u_f}\Bigg\{ a_f(u_f) g(y_f|u_f) \prod_{k\in S_f, k\neq m} p_{jk}(u_f) \\
& \times tr(u_i|u_m,u_f) \\
& \times \prod_{s\in C_{mf}, s\neq i}\Bigg[\sum_{u_s} tr(u_s|u_m,u_f) g(y_s|u_s) \prod_{k\in S_s} p_{sk}(u_s)\Bigg]\Bigg\}\Bigg\}
\end{aligned}
\tag{5.9}
$$

This requires $tr(u_i|u_m,u_f)$, which is the conditional probability that i has genotype u_i given that its parents m and f have genotypes u_m and u_f. This links the joint probability of parental genotypes with joint probabilities of full sib genotypes in Equation 5.9. If no markers are available $tr(u_i|u_m,u_f)$ is one of 0, 1.0, 0.5 or 0.25. If markers are available, $tr(u_i|u_m,u_f)$ is 0, 1.0 , or an intermediate value estimated from the observed marker transmission and the estimated recombination rate.

The equation to calculate posterior probability $p_{ij}(u_i)$ for i with its mate j is:

$$
\begin{aligned}
p_{ij}(u_i) = \quad & \sum_{u_j}\Bigg\{ a_j(u_j) g(y_j|u_j) \prod_{k\in S_j, k\neq i} p_{jk}(u_j) \\
& \times \prod_{o\in C_{ij}}\Bigg[\sum_{u_o} tr(u_o|u_i,u_j) g(y_o|u_o) \prod_{k\in S_o} p_{ok}(u_o)\Bigg]\Bigg\}
\end{aligned}
$$

5.2.3 Comparison of Methods to Calculate Genotype Probabilities

The two methods of estimating genotype probabilities were compared in two different ways to provide useful results for the different uses of genotype probability estimates. These comparisons included use of estimated number of favourable QTL alleles, which was calculated as:

[prob (one allele carried) + 2 x prob (two alleles carried)]

The comparison statistics were:

- Correlation - between true and estimated number of favourable QTL alleles carried
- Ranking on estimated number of favourable QTL alleles carried and based on this ranking selection of top five and 25% of animals in the population

5.2.4 Simulation of Test Data Set

Stochastic simulation was used to evaluate the usefulness of the genotype probability estimates. For each animal a genetic marker locus was simulated linked to a gene of major effect (A_1). Ten distinct marker alleles were simulated in the base generation in linkage equilibrium with the gene of major effect. Simulation consisted of populations of 160 animals with a gene of major effect (A_1) with gene frequency ($p = 0.2$) in the base population. Genotype values a and d were set to 1.02 and 0 respectively (
Table 5.1). The base population consisted of 32 animals, of which 2 sires were selected at random each to be mated to 4 randomly selected dams who produced 4 offspring each. This gave 32 animals per generation for four generations following the base population. Phenotypes were simulated as a function of the true genotype. Recombination rate between the genetic marker locus and QTL was 0.05.

Table 5.1 QTL Genotype classes, frequencies and values (Falconer, 1989).

Genotype	Frequency	Value
A_1A_1	p^2	+a
A_1A_2	2pq	d
A_2A_2	q^2	-a

* $\alpha_1 = q\,(a+d\,(q-p))$, $\alpha_2 = -p\,(a+d\,(q-p))$

Phenotypes for each animal were a function of polygenic and QTL genotype and a normally distributed error term. The base generation animals were simulated according to Equation 5.10.

$$phenotype = TBV_P + TBV_Q + \sqrt{\sigma_e^2} * z \qquad (5.10)$$

where:

TBV_P	is the polygenic true breeding value ($\sqrt{\sigma_u^2} * z_1$)
TBV_Q	is the QTL genotype value (from Table 5.1)
z	is a normally distributed random number $N(0,1)$
z_1	is a second normally distributed random number $N(0,1)$

In the progeny generations the polygenic true breeding value was equal to half the TBV_P of the sire and the dam as well as a Mendelian sampling term. The TBV_Q was a function of the animals inherited QTL genotype.

Given simulated gene frequencies (p and q), genotype values (a, d and - a) and the environmental variance (σ_e^2) the total genetic variance was calculated according to $\sigma_a^2 = 2pq\left[a + d(q-p)\right]^2$. Variance components were then partitioned as follows:

$$\begin{aligned} \sigma_p^2 &= \sigma_a^2 + \sigma_e^2 \\ &= \sigma_u^2 + 2\sigma_v^2 + \sigma_e^2 \\ h^2 &= \sigma_a^2 / \sigma_p^2 \end{aligned}$$

200 replicate populations were simulated for each of two traits, both with a QTL heritability of 0.25, but varying in the proportion of variance due to polygenes. Trait 1 had a small polygenic effect (5% of the total genetic variance) and was considered to be the "No Polygenes Option" and Trait 2 had a polygenic effect of 33% of the total genetic variance, the "Polygenes Option". This gave a heritability of 0.26 for Trait 1 and 0.50 for Trait 2. For both traits analysis was carried out on both selected an unselected populations. Selection was on estimated breeding value calculated using pedigree and phenotype information.

5.3 Results

Results for comparison of the iterative and non-iterative segregation analysis methods are presented below. All results are means of 200 replicate populations. Results in Table 5.2 and Table 5.3 are average correlations between true and estimated number of favourable alleles for simulations involving an initial favourable QTL frequency of 0.2 and 0.5 respectively. Table 5.4 and Table 5.5 give the results from the ranking of animals on the number of favourable alleles they carry when the top 5% of the population is selected.

Table 5.2 Average correlations between true and estimated number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.2

Method of Analysis	**No Selection**				**Selection**			
	No Polygenes		**Polygenes**		**No Polygenes**		**Polygenes**	
	m1[1]	**m10**[2]	**m1**	**m10**	**m1**	**m10**	**m1**	**m10**
Non-iterative	*0.54*[3]	0.59	0.26	0.34	0.75	0.81	0.56	0.66
	(0.01)[4]	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Iterative	*0.56*	0.68	0.34	0.48	0.77	0.86	0.61	0.71
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)

[1]m1 = only one marker allele - ie no marker information available
[2]m10 = 10 marker alleles
[3]values in same columns in italics are not significantly (P< 0.05) different from each other, all other values in same columns are significantly difference from each other
[4]standard errors in brackets

The iterative method (incorporating segregation analysis) gave significantly higher correlations ($P < 0.05$) than the proposed non-iterative method between true and estimated number of favourable alleles in simulations shown in Table 5.2 which had an initial QTL frequency of 0.2, except for the first option. This option involved no selection, no polygenes and no marker information (m1) and showed no significant difference between methods in average correlation. Also evident in Table 5.2 is the fact that correlations were consistently higher for traits influenced by selection, the ability to use marker information and when no polygenic effects were present.

The effect of an increased initial favourable QTL frequency from 0.2 to 0.5 was to increase correlations between true and estimated numbers of QTL alleles in the absence of selection (Table 5.2 and Table 5.3). The correlations were higher with an initial QTL frequency of 0.5, however, when selection then took place the correlations decreased as the frequency become one. Also associated with this increase in initial QTL allele frequency and selection is the fact that the iterative QTL genotype probability estimation method has significantly higher correlations between true and estimated number of favourable QTL alleles than the non-iterative method (Table 5.3). This effect is only present with selection and when polygenes are present, although there is also no difference between methods with selection, when polygenes are absent and with the use of marker information.

Table 5.3 Average correlations between true and estimated number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.5

Method of Analysis	**No Selection**				**Selection**			
	No Polygenes		**Polygenes**		**No Polygenes**		**Polygenes**	
	m1[1]	**m10**[2]	**m1**	**m10**	**m1**	**m10**	**m1**	**m10**
Non-iterative	0.55	0.64	0.28	0.36	0.61	*0.70*[4]	0.50	0.54
	(0.01)[3]	(0.01)	(0.01)	(0.02)	(0.01)	*(0.01)*	(0.01)	(0.01)
Iterative	0.60	0.74	0.41	0.59	0.67	*0.72*	0.46	0.50
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	*(0.01)*	(0.01)	(0.01)

[1]m1 = only one marker allele - ie no marker information available
[2]m10 = 10 marker alleles
[3]standard errors in brackets
[4]values in same columns in italics are not significantly ($P< 0.05$) different from each other, all other values in same columns are significantly difference from each other

The ability of the two genotype probability estimation methods to rank animals on the number

of favourable QTL alleles they carry is shown in Table 5.4 and Table 5.5. There were no significant ($P < 0.05$) differences between the two methods of estimating genotype probabilities in the simulations involving no selection and an initial favourable QTL allele frequency of 0.2 (Table 5.4) although the iterative methods were consistantly higher. The effect of polygenes was to reduce the ability of the methods to correctly rank the animals and the effect of markers was to increase the ability of the methods to rank animals correctly.

The average number of favourable alleles captured in the top 5% of the population in simulations involving an initial favourable QTL allele frequency of 0.2 (Table 5.4) and selection were significantly higher for the iterative method of estimating genotype probabilities than for the non-iterative method, except for the situation involving no polygenic effects and no markers, where there was no difference between the methods.

Table 5.4 Average number of favourable QTL alleles captured in top 5% of population when simulated populations were ranked on the number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.2

Method of Analysis	**No Selection**				**Selection**			
	No Polygenes (12.61)*		**Polygenes** (12.61)*		**No Polygenes** (15.95)*		**Polygenes** (15.57)*	
	M1[1]	**m10**[2]	**m1**	**m10**	**m1**	**m10**	**m1**	**m10**
Non-iterative	*9.11*[3]	*10.28*	*6.54*	*7.87*	*15.20*	15.21	11.92	12.31
	(0.28)[4]	*(0.31)*	*(0.32)*	*(0.35)*	*(0.11)*	(0.11)	(0.24)	(0.22)
Iterative	*9.50*	*10.62*	*7.29*	*8.71*	*15.51*	15.80	13.76	14.69
	(0.29)	*(0.32)*	*(0.34)*	*(0.38)*	*(0.11)*	(0.09)	(0.23)	(0.21)

*true number of favourable QTL alleles in top 5% of population
[1]m1 = only one marker allele - ie no marker information available
[2]m10 = 10 marker alleles
[3]values in same columns in italics are not significantly ($P< 0.05$) different from each other, all other values in same columns are significantly difference from each other
[4]standard errors in brackets

With an initial favourable QTL alleles frequency of 0.5 the iterative genotype probability method always captured more of the favourable alleles than the proposed non-iterative method, indicating greater ability to rank the animals on QTL genotype. Surprisingly, the non-iterative method ranked animals worse when 10 marker alleles were present than when no marker information was available, however, this difference was not significant. Both methods of analysis were able to rank the top 5% of the population almost correctly in the simulations involving selection and no polygenes and with initial QTL allele frequency of both 0.2 and 0.5 (Table 5.4 and Table 5.5). This is evident as both methods captured over 15 favourable QTL alleles when the true number of alleles was 16.

Table 5.5 Average number of favourable QTL alleles captured in top 5% of population when simulated populations were ranked on the number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.5

Method of Analysis	**No Selection**				**Selection**			
	No Polygenes (15.98)*		**Polygenes** (15.98)*		**No Polygenes** (16.00)*		**Polygenes** (16.00)*	
	m1[1]	**m10**[2]	**m1**	**m10**	**m1**	**m10**	**m1**	**m10**
Non-iterative	13.51[3]	14.16	10.28	10.97	15.65	15.39	14.31	13.93
	(0.16)[4]	(0.13)	(0.25)	(0.25)	(0.05)	(0.07)	(0.14)	(0.14)
Iterative	14.08	15.12	12.34	13.94	15.85	15.97	15.23	15.62
	(0.13)	(0.11)	(0.20)	(0.22)	(0.04)	(0.02)	(0.09)	(0.07)

*true number of favourable QTL alleles in top 5% of population
[1]m1 = only one marker allele - ie no marker information available
[2]m10 = 10 marker alleles
[3]all values in same columns are significantly difference from each other
[4]standard errors in brackets

5.4 Discussion

This chapter has shown results from the use of a proposed non-iterative method for estimating genotype probabilities and a segregation analysis method of estimating genotype probabilities. Stochastic simulation was used to test the methods over a range of data sets. Under these simulations good estimates of genotype probabilities were obtained in the absence of polygenic effects, incorporating genetic marker information and in the presence of selection. In one case the correlations were both above 80% (Table 5.2). The effect of each factor contributing to the performance of the genotype probability estimation methods is discussed below.

5.4.1 Effect of Genetic Markers

The addition of genetic marker information allowed higher correlations between true and estimated numbers of favourable QTL alleles to be estimated, as well as improved ranking of animals on the number of favourable QTL alleles they carry. This increase in correlation is caused by the QTL still having a large effect, even though 10 marker alleles were used. That is, the heritability of the QTL effect was at 0.25 for all simulations examined. In the absence of polygenic effects the QTL explained all the additive genetic variance, and even when polygenic effects were added, the QTL effect explained a third of the additive genetic variance. Genetic marker information will be more useful as the variance of the QTL effect as a proportion of the total genetic variance decreases.

5.4.2 Effect of Polygenes

Both methods of estimating genotype probabilities easily encompass both polygenic and QTL effects, as they are both fitted in the mixed model equations given by Equations 1 and 6. Accuracy of estimation of QTL genotype probabilities is reduced as the QTL explains less of the total additive genetic variance, this is shown in the reduced correlations between true and estimated number of favourable QTL alleles (Table 5.2 and Table 5.3).

5.4.3 Effect of Selection

The initial favourable QTL allele frequency was investigated at values of both 0.2 and 0.5. This last value was simulated to investigate the effect of selection. The effect of selection as shown in Table 5.2and Table 5.3 was to increase the correlations between true and estimated numbers of favourable alleles. It was unsure as to whether this was simply caused by the increase in gene frequency of the favourable QTL allele when selection was present or if it was caused by the ability of the methods of estimating genotype probabilities to account for selection. With an initial favourable QTL allele frequency of 0.5 and selection the frequency of favourable alleles will become one, making it harder to detect differences between animals. However, the increase in initial allele frequency was able to allow for more accurate ranking of animals in the top 5% of the population in the presence of Table 5.4 and Table 5.5)

5.4.4 Effect of Method of Analysis

Comparisons of correlations between true and estimated number of favourable alleles carried by animals within each simulated pedigree show that the proposed non-iterative genotype probability estimation method does not provide as accurate estimates as the segregation analysis method (except in the case of both polygenic effects and selection). The latter mentioned situation is slightly problematic, as the interpretation is unclear. It is unknown why the proposed non-iterative method would produce significantly greater correlations than the iterative method under these conditions. The non-iterative method had much higher selection response under these conditions indicating that as the gene frequency went up there might have been a negative affect on the correlation. There are also significant differences ($P < 0.05$) between the methods tested when animals within each pedigree are ranked on the number of favourable alleles except with low initial gene frequency and in the absence of selection. This provides for the situation where the proposed method would be the most use. That is, for a population that has not undergone selection and at a low initial gene frequency the proposed non-iterative method of estimating genotype probabilities will be nearly as useful as the iterative method for ranking animals.

Adequacy of non-iterative genotype probability estimation method

The iterative method of estimating genotype probabilities more correctly models the true situation of two QTL alleles segregating. That is, the model estimated by the iterative method is the same as the model simulated. The proposed non-iterative method assumes a normal density for the QTL of interest with a given gene frequency, variance of QTL effect and mean for each genotype. This raises questions as to the adequacy of the latter model as suggested by Guo and Thompson, (1992). However, in real life there may be more variation in QTL effects due to interactions with background genes. This would suggest that the simulated model was more in favour of the fixed (or iterative) model, although the non-iterative method also assumes equal QTL effects. The greatest attraction of the non-iterative approach continues to be its ability to handle complex pedigrees as easily as it does simple pedigrees. It is, therefore, conceptually simple and computationally feasible.

Application of the proposed procedure to estimate genotype probabilities requires estimates of prior gene frequency and genotypic means. It assumes that the gene frequencies in the base population are in Hardy Weinburg equilibrium. However, if better estimates of the gene frequencies are available they are easily accommodated in the probability estimation. Knowledge of recombination rates between QTL and linked marker loci are required, as is the additive genetic variance due to the polygenic and marked QTL effects. However, these could be estimated from the data during linkage analysis prior to estimation of breeding values and genotype probabilities.

Test of iterative genotype probability estimation method

The effectiveness of the iterative method of estimating genotype probabilities was not affected by the small and looped pedigree. With only four generations of animals born after the base generation and only two males selected as sires each year the pedigree became inbred very quickly. Although, slowed down by increased complexity pedigrees (i.e slower speed of convergence), the method presented for estimating genotype probabilities is relatively robust over a range of situations including both with and without marker information and for traits both controlled by polygenes and single QTL effects. The application of the method over a less looped and inbred pedigree should be to improve its performance, indicating it to be a relatively robust method for calculation of genotype probabilities over a range of pedigree types.

5.5 Conclusions

The non-iterative method present in this chapter is based on the model of Fernando and Grossman (1989) which can be used to estimate breeding values based on major genes as well as polygenic effects. Genotype probabilities are derived from estimated random major gene effects by simple extension to the method. Besides being easier to apply, this method also circumvents the problem of having to deal with looped pedigrees. This study has shown that genotype probabilities from this procedure are not as accurate as an iterative procedure where polygenic, major gene effects and genotype probabilities are estimated conditional on each other. However, other genetic models may have affected the comparison, such as the simulation of fixed QTL effects which could favour the iterative method estimation.

Chapter 6

Use of Marker Information and Genotype Probabilities in Mate Allocation and Selection Decisions

6.1 Introduction

In previous chapters, the use of marker and QTL information was examined for use in linkage detection and breeding value estimation. Use of marker assisted selection was assumed to be of most benefit to animal breeders in estimating breeding values (EBVs) for individual QTL alleles. This would be the case if QTL alleles were found that affected traits of primary importance and which explained a large amount of the additive genetic variance. Of interest in this chapter is an alternative form of marker assisted selection, where marker and major gene information is used in *genetic value estimation* (Kinghorn and Clarke, 1996). Genotype probabilities may be used in selection of mating pairs. Mate allocation could then be optimised with respect to expected progeny merit (in this case, expected progeny genotype).

In the case of selection based on EBV (of one trait, or a combination of traits in an index), expected progeny merit is the mid-parent value of mating pairs. This assumes that inheritance is additive and the combination of additive genetic effects within the progeny is linear (Bunter, 1995). In situations where there is opportunity to predict genotypes and exploit interaction directly, alternative approaches are required to maximise progeny merit.

Predicting progeny genotypic value is generally important if dominance plays a role in the phenotypic expression of a trait. An extreme example of this is where unfavourable recessive alleles are segregating that affect traits which are not normally in the breeding objective. Examples are lethal alleles, and genes coding for diseases, genetic defects or other undesirable traits like horns. The frequency of these alleles in the population may be minimised, or alternatively, the expression of the homozygous carrier genotype may be minimised, by including estimated genotype probabilities in mate selection decisions. Spending selection effort in controlling such traits normally results in loss of genetic gain for the traits of primary importance in the breeding objective.

Non-random mating, as described by Allaire (1977), allows potential mates to be ranked by the EBV of their expected progeny, instead of by the EBV of each candidate parent. The aim is to develop a mating method that achieves a higher total progeny merit than is possible with random mating among selected parents. As stated by Allaire and Barr (1990), "In general, the mating of the best bull and the best cow will not correspond to the best mating pair unless the total merit of their progenies arises as the strictly additive contribution from their parents' genetic values". For multi-trait selection, index values for expected progeny among potential mates should be used when a non-linear relationship exists between at least one-trait in the index and merit (Allaire, 1980; Allaire *et al.*, 1985). Linear programming techniques are optimal methods for solving mate allocation problems where evaluating all mating combinations is not feasible for more than a few candidates (Jansen and Wilton, 1985; Kinghorn, 1987). Linear programming considers all possibilities implicitly while only explicitly evaluating a small subset of them.

In Merino sheep breeding, the traits of primary economic importance include fibre diameter and fleece weight. Traditionally selection is carried out to maximise fleece weights and minimise fibre diameter. However, there are traits which are not selected for in the breeding objective, but that cause a loss of profit if no selection effort is put on them. Examples of such traits include pigmented fibres and the recessive allele for black lambs. Although not lethal in the strict sense, the latter recessive allele may be thought of as lethal, as expression of the condition causes animals to be automatically culled from the breeding flock. Culling animals for such traits results in a loss of selection intensity and therefore a loss of selection response for the traits of primary economic importance.

This extreme culling procedure is due to the financial losses that occur for individual wool processors and/or fabric/garment manufacturers when dark fibres contaminate white wool (Fleet, 1996). For all white sheep flocks, the danger with coloured sheep is their potential to transfer pigmented fibres to white sheep in paddocks or yards and to white wool being prepared for market in shearing sheds. As such, wool quality management systems require the removal of coloured or partly coloured animals from the property. The problem in ram-breeding flocks is that the use of a carrier ram can lead to a rapid increase in the frequency of the undesired allele and dissatisfied customers. It is thought that on average 6% of white Merino sheep are carriers of the recessive allele for black lambs (Fleet *et al.*, 1995). At present the only practical means of determining whether a ram carries this recessive allele is by progeny testing, although work is proceeding to identify carriers based on molecular testing of DNA (Fleet *et al.*, 1995; Parsons *et al.*, 1997a).

This study examines different strategies for including marker information on lethal genes in the breeding program by using mate selection. The value of marker information as well as the impact of using mate selection strategies was assessed in terms of reducing the frequency of black lambs and genetic gain for fibre diameter (FD). A closed nucleus flock was simulated, with no new alleles able to enter the flock. This scenario was chosen as it is only the stud sector which can afford to genotype animals for the black gene for pigmented fibre, which was assumed to be segregating in the population.

Different strategies for selection of rams and ewes were chosen to examine the effects of the following options on the ability to make gains in FD while reducing the incidence of black lambs in the breeding flock: 1) an efficient (but costly) genotyping and mate selection strategy, 2) a lower cost selection strategy using genotype information from males only and mate selection and 3) a minimal cost selection strategy using only phenotypic information, genotype probabilities and mate selection. A comparison of these strategies was made against reference strategies: traditional phenotypic selection against black lambs combined with selection for FD EBVs and single trait selection for either FD EBV or against black lambs.

The aim of the study was, therefore, to determine the possible increase of selection response against deleterious alleles by using genotypic information (or linked marker information) while maximising gain for trait of interest (FD)

6.2 Materials and Methods

6.2.1 Simulation Procedures

A stochastic simulation was designed for a population structure with a fixed number of dams per sire and fixed total number of dams mated to selected sires each generation. This population represented a closed nucleus stud Merino breeding flock with a number of age classes, where continuous evaluation for selection, mating and progeny generation is carried out. Consequently, the population had overlapping generations, constrained by the number of years sires and dams were available as parents. Breeding values were simulated for fibre diameter and genotypes were simulated for the black wool gene.

6.2.1.1 Parameters for Simulated Flock

The closed breeding flock underwent selection for 10 years following simulation of the base animals. The base population consisted of 200 animals (100 males and 100 females). Two subsequent generations were simulated. This involved 2 years of random mating of 5 randomly selected sires to 100 randomly selected dams, each of whom produced either 1 or 2 progeny, giving a total of 150 progeny per year. It was assumed that animals could be used as parents in the five years following their birth, and for simplicity there was no mortality over years.

Parameter values defining the simulated flock are included in

Table 6.1. After three generations a total of 500 animals were created. 150 new animals were simulated from matings of selected sires and dams in each of the following 10 years. This gave a total of 2000 animals with records. Animals not selected as parents were not available as parents in the following years. In the given flock structure, 5 males were selected from approximately 80 males to be sires for the next mating round and 100 females were selected from approximately 175 females to be dams each year.

Table 6.1 Parameter values used for flock simulation

Restriction Parameters	Values
Number of dams per sire for yearly matings	10
Total number of dams used for yearly matings	100
Number of progeny per dam	1.5
Number of years sires available as parents	2
Number of years dams available as parents	5
Number of years of selection	10

6.2.1.2 Simulation of Records

All animals (except those from the first year of simulation) had full pedigree, sex, phenotypic and genotype information generated. Animals in the first year of simulation had no parent records. Phenotypic observations were recorded for fibre diameter measurements and expression of black wool.

Fibre Diameter (FD)

FD phenotypes were assumed to be taken at hogget shearing. Base animals had phenotypes simulated according to the values in Table 6.2, the mean and variance being Australian industry estimates (K.D. Atkins, *pers comm*) for 12 -15 month hoggets. The phenotype (y) for each base animal (i) was simulated as the sum of the mean (μ), the true breeding value (TBV_i) and error:

$$y_i = \mu + TBV_i + \sqrt{\sigma_e^2} * z_1$$

Where:

$$TBV_i = \sqrt{\sigma_a^2} * z_2$$

z_1 and z_2 are normally distributed random variables $N(0,1)$

The true breeding value of each progeny (p) was calculated as the mean of the true breeding values of its sire (s) and dam (d):

$$TBV_p = \frac{TBV_s + TBV_d}{2} + \sqrt{0.5 * \sigma_a^2} * z_3$$

where :z_3 is a normally distributed random variable $N(0,1)$ and the effect of inbreeding on variance was ignored.

Table 6.2 Fibre diameter parameters used for phenotypic simulations

Parameter	Value
Mean (μ)	21 microns
Heritability (h^2)	0.50
Genetic Variance (σ_a^2)	0.9316
Phenotypic Variance (σ_p^2)	1.863
Error Variance (σ_e^2)	0.9316

Black Wool Gene (b)

The black wool allele is recessive and assumed to be uncorrelated with FD for these simulations. Genotyping for the black wool allele (*b*) was assumed to be available at hogget age when phenotypic records for FD were taken. Phenotype for black lambs (*bb*) is known at birth. Animals were recorded as homozygous white (*BB*), heterozygous carriers (*Bb*) or homozygous black (*bb*). The black gene was simulated in the base population at an initial frequency of 0.4. This gave approximately 36% *BB*, 48% *Bb* and 16% *bb* animals before selection and mate allocation. Progeny genctypes were simulated according to transmission probabilities of sires and dams genotypes (Table 6.3).

Table 6.3 Transmission probabilities of inheriting QTL alleles from sires and dams (assuming Mendelian inheritance).

Genotype		**Probabilities of Progeny Genotypes**		
Sire	Dam	BB	Bb	bb
BB	BB	1	0	0
Bb	BB	0.5	0.5	0
bb	BB	0	1	0
BB	Bb	0.5	0.5	0
Bb	Bb	0.25	0.5	0.25
bb	Bb	0	0.5	0.5
BB	bb	0	1	0
Bb	bb	0	0.5	0.5
bb	bb	0	0	1

6.2.1.3 Calculation of Selection Criterion

Selection criteria were Estimated Breeding Value (EBV), derived using BLUP procedures, for fibre diameter (FD), and/or expected genotype probability of progeny, derived from genotype information on potential parents, for the black wool gene. BLUP procedures were carried out using an animal model as described below.

BLUP Selection

EBVs were calculated for FD of each animal using an animal model (Henderson, 1975), incorporating all information from relatives. The only fixed effect fitted was the mean.

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where:

$\mathbf{y}$	is the vector of phenotypic observations
β	is the vector of fixed effects
$\mathbf{u}$	is the vector of random genetic effects
$\mathbf{e}$	is the vector of random residual effects
$\mathbf{X}$ and $\mathbf{Z}$	are known incidence matrices

The variance-covariance structure of random variables is:

$$\mathbf{V}\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_u^2 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix}.$$

where

$\mathbf{A}$	is the numerator relationship matrix
$\mathbf{I}$	is an identity matrix

This gives the following mixed model equations to solve for fixed effects ($\hat{\beta}$) and estimated breeding value ($\hat{u}$):

$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\sigma_e^2/\sigma_a^2\end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X'y}\\ \mathbf{Z'y}\end{bmatrix}$$

The expected breeding value for future progeny (p) of sire (i) and dam (j) was calculated as:

$$EBV_p = \frac{\hat{u}_i + \hat{u}_j}{2}$$

Genotype Probabilities for the Black Gene

The probabilities of sire (*i*) and dam (*j*) producing progeny being *BB*, *Bb* or *bb* are given in Table 6.3. Only sire and dam information was used for calculating expected progeny genotypes, no other pedigree information was used. These probabilities could be calculated for each available mating pair, and were of the form:

prob0	is the probability of producing progeny carrying no *b* allele
prob1	is the probability of producing progeny carrying one *b* allele
prob2	is the probability of producing progeny carrying two *b* alleles

6.2.2 Selection Alternatives

6.2.2.1 Phenotypic Selection Against Black Lambs

A simulated population was selected for FD EBV over 10 years as a control population. This selection was termed "Phenotypic Selection" (**PHS**) as it represented phenotypic selection against black lambs combined with selection for FD EBV. Black animals were unavailable for selection as parents and parents were selected on FD EBV and then randomly mated. This reflects current practice in sheep breeding flocks.

6.2.2.2 Selecting Mating Pairs using Linear Programming - all animals genotyped

Linear programming techniques were used to select mating pairs. This involved simultaneous selection and mating of males and females according to predicted progeny merit. This was possible using an implementation of the "transportation problem" described in Jansen and Wilton (1985) and involved the following two-steps:

(A) Develop an objective function which describes net economic gain as a function of selection and mate allocations
(B) Implement a mate selection algorithm which maximises the objective function

Development of Objective Function

Within each round of selection, we want to maximise a function of EBVs for a trait of economic importance and minimise probabilities of expression of a lethal trait. The objective function in this study was a paired merit function ($f_{ij}(PM)$) comprising both expected progeny breeding value for fibre diameter (EBV_{ij}), multiplied by a weighting factor, and probability of progeny genotype (P_{ij} *or* $\hat{P_{ij}}$) expressing the black condition.

A number of different objectives were evaluated ranging from all emphasis on fibre diameter gain (**AS5**) and none on black wool, to no emphasis on fibre diameter (**AS1**). These objectives are shown in Table 6.4. The weighting's of 0, 50 and 100 were arbitrarily chosen, as there is little information available about the economic value of reducing the incidence of black wool in the Australian wool clip. Also, the expected breeding value of the progeny is multiplied by 50 (in Equation 6.1) to bring the genotype probabilities and fibre diameter estimated breeding values to approximately the same scale, with values sufficiently large to give good resolution in the integer optimisation method used. Two objectives were considered within the mate selection alternatives (Equations 6.2 and 6.3), these involved comparing minimisation of *b* allele frequency (P_{ij}) or minimisation of black lamb expression *bb* ($\hat{P_{ij}}$). The difference between P_{ij} and $\hat{P_{ij}}$ is that P_{ij} minimises the frequency of the *b* allele in the next generation, and $\hat{P_{ij}}$ minimises the expression of *bb* animals, therefore not penalising heterozygous carriers of the *b* allele. Minimising frequency of *b* alleles (P_{ij}) is linear, therefore, random mating of selected parents is optimal. Minimising expression ($\hat{P_{ij}}$) is not linear and requires mate allocation to get optimum gains in the next generation. This has no effect in mating option **AS5** so this option was not repeated for $\hat{P_{ij}}$.

Table 6.4 Paired merit function ($f_{ij}(PM)$) for sire *i* and dam *j*, based on measure of estimated breeding value EBV_{ij} and *b* genotype probability P_{ij} for five mate selection schemes (AS1-AS5). [Paired Merit $f_{ij}(PM)^{\wedge}$ is similar to $f_{ij}(PM)$, except *bb* genotype probabilities are minimised and four mate selection schemes are used (AS1*-AS4*)]

Mating Option Code	Paired Merit ($f_{ij}(PM)$)	Mating Option Code	Paired Merit ($f_{ij}(PM)^{\wedge}$)
AS1	$-100 * P_{ij}$	AS1*	$-100 * P_{ij}^{\wedge}$
AS2	$-100 * P_{ij} - 50 * EBV_{ij}$	AS2*	$-100 * P_{ij}^{\wedge} - 50 * EBV_{ij}$
AS3	$-100 * P_{ij} - 100 * EBV_{ij}$	AS3*	$-100 * P_{ij}^{\wedge} - 100 * EBV_{ij}$
AS4	$-50 * P_{ij} - 100 * EBV_{ij}$	AS4*	$-50 * P_{ij}^{\wedge} - 100 * EBV_{ij}$
AS5	$-100 * EBV_{ij}$		

$$EBV_{ij} = 50 * (EBV_s + EBV_d)/2 \quad (6.1)$$

$$P_{ij} = -100 * prob0 - 50 * prob1 - 0 * prob2 \quad (6.2)$$

$$P_{ij}^{\wedge} = -100 * prob0 - 100 * prob1 - 0 * prob2 \quad (6.3)$$

6.2.2.3 Selecting Mating Pairs using Linear Programming - only males genotyped

In addition to the mating schemes given in Table 6.4, these same options were repeated with only male animals genotyped for the black wool gene. This allowed comparison of earlier mentioned selection options with those involving reduced costs. With only males genotyped, transmission probabilities for use in calculating the paired merit function were different to those given in Table 6.3. Genotypes of the black wool ewes (*bb*) were known, even though animals were not genotyped, as the gene was expressed. The transmission probabilities used for only male genotyping are given in Table 6.5, again only sire and dam information was used in the calculation of genotype probabilities, no other pedigree information was assumed available. Mating option codes were the same as those for all animals genotyped but had A replaced by M (**MS1, MS2, MS3, MS4, MS1*, MS2*, MS3*, MS4***).

Table 6.5 Transmission probabilities of inheriting QTL alleles from genotyped sires and ungenotyped dams (assuming Mendelian inheritance).

Genotype		**Probabilities of Progeny Genotypes**		
Sire	Dam	BB	Bb	bb
BB	BB or Bb	0.75	0.25	0
Bb	BB or Bb	0.375	0.5	0.125
bb	BB or Bb	0	0.75	0.25
BB	bb	0	1	0
Bb	bb	0	0.5	0.5
bb	bb	0	0	1

6.2.2.4 Selecting Mating Pairs using Linear Programming - using genotype probabilities

The final selection options studied examined the use of genotype probabilities instead of known genotype information for the black wool gene (*b*). Segregation analysis (Kerr and Kinghorn, 1996) was used to calculate the probability of a potential parent being in each of the three genotype classes (*BB, Bb* or *bb*). This required only pedigree and phenotype information, where phenotypes were 0 for white sheep and 1 for black sheep. These probabilities were then used to calculate probabilities of progeny genotypes (*prob0, prob1* and *prob2*) using the following equations:

$$prob0 = (s0 * d0) + 0.5(s1 * d0) + 0.5(s0 * d1) + 0.25(s1 * d1)$$

$$prob1 = 0.5(s1 * d0) + (s2 * d0) + 0.5(s0 * d1) + 0.5(s1 * d1) + 0.5(s2 * d1) + (s0 * d2) + 0.5(s1 * d2)$$

$$prob2 = 0.25(s1 * d1) + 0.5(s2 * d1) + 0.5(s1 * d2) + (s2 * d2)$$

where: *s0, s1* and *s2* are the probabilities of the sire having 0, 1 or 2 copies of the black gene (*b*) respectively, and *d0, d1* and *d2* are the probabilities of the dam having 0, 1 or 2 copies of the black gene (*b*) respectively. Again, the mating option codes were the same as in Table 6.4, except they had the prefix **P** representing use of genotype probabilities (**PS1, PS2, PS3, PS4, PS1*, PS2*, PS3*,PS4***).

6.2.3 Mate Selection Procedures using Linear Programming

Mate selection is used for **AS**, **MS** and **PS** selection options. The theory behind use of linear programming to select mating pairs as used for these selection options is described below. Within each round of selection, we want to maximise response in the next generation according the objectives defined in Table 6.4. For linear programming, this involves minimising a cost function (c) for all potential mating pairs of sires i and dams j. In linear programming terminology the problem may be formulated as minimising the matrix of costs c_{ij} ($i = 1,...m$, $j = 1, ...n$) times the matrix of number of transportations. The costs in this study are each of the paired merit functions given in Table 6.4. This matrix is the cost of transporting (mating) a unit from the *i'th* source (dam) to the *j'th* supply (sire) and minimising this cost is subject to the constraints that:

a) females are mated at most once
b) the *i'th* male is mated no more than k_i times
c) the maximum number of mating pairs is f

From Jansen and Wilton (1985) this may be expressed as:
The aim is to minimise:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}X_{ij} \tag{6.4}$$

subject to the constraints:

$$X_{ij} = 1 \quad or \quad 0 \tag{6.5}$$

$$\sum_{i=1}^{m} X_{ij} \leq 1 \quad for \quad j = 1,n \tag{6.6}$$

$$\sum_{j=1}^{n} X_{ij} \leq k_i \quad for \quad i = 1,m \tag{6.7}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} = f \tag{6.8}$$

where X_{ij} is the number of progeny from the ij^{th} mating pair. Including a dummy male and female allows constraints (6.6) and (6.7) to become equalities:

$$\sum_{i=1}^{m} X_{ij} = 1 \quad for \quad j = 1,n$$

$$\sum_{j=1}^{n} X_{ij} = k_i \quad for \quad i = 1,m$$

and additional constraints for the dummy animals are:

$$\sum_{i=1}^{m+1} X_{i,n+1} = \sum_{i=1}^{m} k_i - f$$

$$\sum_{j=1}^{n+1} X_{m+1,j} = n - f$$

For (6.8) to hold, $X_{m+1,n+1}$ must equal 0 in the optimum solution. This is done by setting the cost function of all dummy matings equal to each other and less than minimum merit for real matings.

The linear program used requires negative integer input values, hence the multiplication of FD EBV by 50 (Equation 6.1). The value of dummy matings, as described above, was set to -999,999, therefore all values of paired merit (or cost) had to be negative, but no smaller than -999,999. The costs functions that were minimised were integers of the paired merit functions described in Table 6.4.

6.2.4 Summary Statistics

Following ten years of mate selection, the following summary statistics were recorded for the 500 base animals and the 150 animals born each year (i.e. before selection):

p(b)	the percent of animals carrying 1 or 2 copies of the *b* allele
p(bb)	the percent of *bb* animals
mfd	the mean FD
mtbv	the mean true breeding value for FD

These summary values were averaged over the 100 replicates of all selection options simulated, and were used to examine response to selection.

6.3 Results

In this section, results from simulated selection and mating schemes are presented. Values represent means from 100 replicate simulations, means for each year represent the mean of all animals born in that year. Results are split into four options which can be summarised as:

1. Reference Selection Options: These include the phenotypic selection against black wool and selection for FD EBV (**PHS**) and single trait selection against the *b* allele (**AS1, MS1, PS1**), against black lambs (**AS1*, MS1*, PS1***) and for FD EBV (**AS5**). Extremes are selection against black wool (**1**) to ignoring black wool (**5**), with other suffixes for intermediate values.
2. High Cost Selection Options: This includes options which involve genotyping all animals (**AS2, AS3, AS4, AS2*, AS3*** and **AS4***)
3. Medium Cost Selection Options: This includes options which involve only genotyping males (**MS2, MS3, MS4, MS2*, MS3*** and **MS4***)
4. Low Cost Selection Options: These options do not involve any genotyping, instead genotype probabilities are calculated for each animal (**PS2, PS3, PS4, PS2*, PS3*** and **PS4***)

6.3.1 Reference Selection Options

Results from phenotypic selection against black lambs and selection for FD EBVs (**PHS**) are given in Table 6.6. There was a steady decrease in numbers of animals carrying *b* alleles. Both heterozygotes and homozygous black lambs decreased in numbers over the 10 year selection period. This was due to black animals not being available as parents in the next generation. There were significant gains in FD made over the 10 years, with a decrease from 21 to 15 microns.

Results from simulation options **AS1** (**AS1***) and **AS5** gives maximum possible gains over the 10 year period of selection for black genes and FD respectively (Table 6.6). With 100 replicate populations, the proportion of 500 animals in each genotype class prior to selection was 36% *BB*, 48% *Bb* and 16% *bb* as expected. The aim of the **AS1** selection option was to minimise the frequency of the *b* allele from the population, and the simulation results show

that it was possible to remove the *b* allele from the population with three years of selection and mate allocation. No black lambs were born from the first year onwards. The **AS1*** option was concerned with simply minimising the expression of *bb* animals in the population, hence there was no penalty for having heterozygous animals. Through efficient mate allocation, no black lambs were born after the first generation, even though there were from 28.7 – 56.8% *b* allele carriers in the flock over a 10 year period.

The **AS5** option involved no selection against the black wool gene, this option only included estimated breeding values for FD. Following 10 years of selection, there were approximately equal proportions of all 3 black gene genotypes (Table 6.6).

Table 6.6 Results of selection on EBV for FD and random mating with black animals (*bb*) not able to be parents (**PHS**), selection for genotype probabilities, P_{ij} (**AS1**) and $\hat{P}_{ij}$ (**AS1***) and estimated breeding value for fibre diameter EBV_{ij} (**AS5**)

Year	**PHS**			**AS1**			**AS1***			**AS5**		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02	64.00	15.88	21.02
1	48.73	8.15	20.18	18.92	0.00	21.03	56.83	0.00	21.06	64.73	16.57	20.01
2	41.95	5.81	19.54	0.44	0.00	21.01	44.79	0.00	21.03	63.46	16.63	19.38
3	36.02	4.70	18.98	0.00	0.00	21.01	39.63	0.00	21.02	62.15	17.83	18.75
4	33.08	3.97	18.42	0.00	0.00	21.01	37.52	0.00	21.04	61.66	18.44	18.17
5	30.34	3.79	17.85	0.00	0.00	21.06	34.49	0.00	21.07	61.71	20.70	17.57
6	26.12	3.14	17.29	0.00	0.00	21.08	33.34	0.00	21.10	61.82	22.92	16.99
7	26.44	3.14	16.72	0.00	0.00	21.11	31.35	0.00	21.11	60.45	23.97	16.40
8	23.31	2.55	16.16	0.00	0.00	21.16	29.71	0.00	21.15	59.39	24.73	15.80
9	22.49	2.80	15.62	0.00	0.00	21.21	29.59	0.00	21.20	59.29	26.85	15.22
10	21.59	2.69	15.04	0.00	0.00	21.27	28.73	0.00	21.24	59.20	27.43	14.60

[1] *p(b)* mean proportion of *b* allele carriers
[2] *p(bb)* mean proportion of black animals
[3] *mfd* mean FD

FD was not selected for in options **AS1** and **AS1***, hence there was no decrease in FD over time with these options (Table 6.6). Option **AS5** showed the maximum possible gains in FD over the 10 year period, this was a decrease from 21.02 microns to 14.7 microns. This is quite a large decrease and does not take account of any fixed effect factors which affect FD. However, option **PHS** was not much worse, showing a final FD of 15.04.

The effect of only genotyping males alone was to lengthen the time over which the *b* allele could be removed from the population. In fact, with selection on nothing but P_{ij} (option **MS1**), the *b* allele could not be completely removed from the flock over the 10 year selection period (Table 6.7). The difference between genotyping all animals and only genotyping males was dramatic for selection options involving only minimising expression of black lambs (**AS1*** and **MS1***) was dramatic. When only males were genotyped, not only was the expression of the black genotype reduced compared to **AS1**, but there was also a decrease in the proportion of heterozygote *b* allele carrier animals. This is due to unknown genotypes of dams requiring all males of the *BB* genotype to be mated to either *BB* or *Bb* females. Over time, this results in reduced frequency of heterozygous progeny. When all animals were genotyped heterozygous sires could be mated to *BB* dams with no resultant *bb* progeny. Selection on genotype probabilitites against the black gene only had no effect on FD gains, so these results are not shown here.

The effect of not genotyping any animals, and using segregation analysis to calculate genotype probabilities for the black wool gene, was that the time required to remove the gene from the population was increased. Even in the selection option involving removal of the gene completely (**PS1**) the gene was not able to be completely removed over the 10 year period (Table 6.7), and a proportion of black lambs continued to be born in the flock up until year 9. The proportion of black lambs born was actually slightly reduced in the selection option which allowed heterozygote *b* allele carriers (**PS1***). However, in both these options there was less than 1% of black lambs born after year 1 (Table 6.7).

Table 6.7 Results of selection for genotype probabilities with only males genotyped, P_{ij} (**MS1**) and $P_{ij}^{\wedge}$ (**MS1***) and with no animals genotyped (using segregation analysis) P_{ij} (**PS1**) and $P_{ij}^{\wedge}$ (**PS1***)

Year	MS1		MS1*		PS1		PS1*	
	$p(b)$[1]	$p(bb)$[2]	$p(b)$	$p(bb)$	$p(b)$	$p(bb)$	$p(b)$	$p(bb)$
0	64.27	16.37	64.27	16.37	64.27	16.37	64.27	16.37
1	29.24	0.00	41.31	0.00	25.15	1.18	26.30	1.24
2	18.51	0.00	25.47	0.00	13.37	0.41	14.05	0.31
3	11.19	0.00	14.59	0.00	7.24	0.15	7.37	0.17
4	6.55	0.00	8.69	0.00	2.99	0.00	3.48	0.03
5	4.07	0.00	5.49	0.00	1.56	0.01	1.94	0.03
6	2.71	0.00	3.22	0.00	0.81	0.02	0.95	0.00
7	1.85	0.00	1.92	0.00	0.70	0.02	0.99	0.01
8	1.31	0.00	1.23	0.00	0.18	0.00	0.80	0.01
9	0.69	0.00	0.79	0.00	0.18	0.01	0.67	0.00
10	0.51	0.00	0.58	0.00	0.07	0.00	0.58	0.00

[1] $p(b)$ mean proportion of *b* allele carriers
[2] $p(bb)$ mean proportion of black animals

6.3.2 High Cost Selection Options

An index involving both equal and unequal weighting on the 2 selection criteria P_{ij} and EBV_{ij} (and $P_{ij}^{\wedge}$ and EBV_{ij}) was used for selection options **AS2**, **AS3** and **AS4** (and **AS2***, **AS3*** and **AS4***). Resultant proportions of *b* allele carriers and black lambs for selection options **AS2**, **AS3** and **AS4** are shown in Table 6.8. There was greater weighting on P_{ij} in option **AS2**. This produced similar results to the maximum gains in option **AS1**, with no black lambs being born after generation one, and removal of the *b* allele completely after four years of selection (which was one more year than with option **AS1**). Selection option **AS4** had greater weighting on EBV_{ij} than P_{ij}, this resulted in the *b* allele not being removed from the population after 10 generations of selection. Heterozygote carrier animals remained in the population throughout the selection period, and black lambs continued to be born into the flock. With approximately equal weighting on P_{ij} and EBV_{ij}, option **AS3** was able to remove the *b* allele completely in nine years of selection, with no black lambs being born after five

years of selection.

Table 6.8 Results of combined selection on genotype probabilities aimed at minimising frequency of *b* allele (P_{ij}) and estimated breeding value for fibre diameter (EBV_{ij}) with all animals genotyped

Year	**AS2**[a]			**AS3**[b]			**AS4**[c]		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02
1	19.59	0.06	20.40	27.82	1.46	20.27	41.42	5.39	20.16
2	2.01	0.00	19.83	11.58	0.13	19.68	27.22	2.47	19.52
3	0.14	0.00	19.19	4.14	0.01	19.09	19.14	1.37	18.94
4	0.00	0.00	18.59	1.57	0.01	18.50	13.04	0.66	18.34
5	0.00	0.00	17.99	0.36	0.00	17.90	7.85	0.23	17.77
6	0.00	0.00	17.38	0.14	0.00	17.31	4.86	0.15	17.20
7	0.00	0.00	16.78	0.03	0.00	16.76	3.30	0.13	16.62
8	0.00	0.00	16.18	0.01	0.00	16.16	2.58	0.19	16.00
9	0.00	0.00	15.59	0.00	0.00	15.57	1.78	0.09	15.43
10	0.00	0.00	15.01	0.00	0.00	14.96	1.22	0.11	14.85

[a]weight on P_{ij} : weight on EBV_{ij} = 2:1
[b]weight on P_{ij} : weight on EBV_{ij} = 1:1
[c]weight on P_{ij} : weight on EBV_{ij} = 1:2
[1]*p(b)* mean proportion of *b* allele carriers
[2]*p(bb)* mean proportion of black animals
[3]*mfd* mean FD

In contrast, results of selection on EBV_{ij} for FD gains (Table 6.8), show that option **AS4** was able to obtain greater response than option **AS3** (which had greater weighting on P_{ij}). This is most likely caused by option **AS3** putting too much selection effort on completely removing the *b* allele in nine generations, losing selection intensity for selection on FD gains. Option **AS2**, however, removed the gene quickly (in four years), but genetic gain in FD suffered greater loss than in options **AS4** and **AS3**.

With the aim of minimising expression of black lambs, option **AS2*** (with greater weighting on P_{ij}^ than EBV_{ij}), produced genotype results similar to option **AS1*** (which did not include selection for EBV_{ij}). The differences between all options involving minimising expression of black lambs (**AS2***, **AS3*** and **AS4***) were that they took 1, 2 and 7 years of selection to

produce no black lambs (Table 6.9). This form of selection against black lambs did not have a negative effect on FD selection, producing similar improvements in FD to 14.8 and 14.72 microns. These were not very different from the maximum gains possible selecting on EBV_{ij} only (option **AS5**) which was 14.7 microns.

Table 6.9 Results of combined selection on genotype probabilities aimed at minimising black lambs ($P_{ij}^{\wedge}$) and estimated breeding value for fibre diameter (EBV_{ij}) with all animals genotyped

Year	**AS2***[a]			**AS3***[b]			**AS4***[c]		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02
1	60.31	0.00	20.12	61.31	0.48	20.12	63.01	1.86	20.10
2	52.54	0.00	19.48	52.68	0.17	19.47	54.82	1.50	19.45
3	46.71	0.00	18.88	47.47	0.00	18.86	49.58	0.79	18.88
4	43.79	0.00	18.28	44.33	0.00	18.26	46.10	0.09	18.29
5	41.35	0.00	17.69	42.65	0.00	17.66	43.00	0.00	17.70
6	38.33	0.00	17.13	40.26	0.00	17.08	41.52	0.01	17.11
7	36.09	0.00	16.56	38.73	0.00	16.49	38.82	0.01	16.56
8	34.63	0.00	16.00	36.44	0.00	15.89	36.73	0.00	15.96
9	33.54	0.00	15.40	35.22	0.00	15.30	34.65	0.00	15.38
10	32.06	0.00	14.80	35.45	0.00	14.72	33.37	0.00	14.80

[a]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 2:1
[b]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 1:1
[c]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 1:2
[1]*p(b)* mean proportion of *b* allele carriers
[2]*p(bb)* mean proportion of black animals
[3]*mfd* mean FD

The differences in true breeding values (TBVs) between all methods involving genotyping all animals and the maximum gains possible using FD as the only selection criteria are presented in Figure 6.1. These results show that the methods involving minimising expression of black lambs (**AS2***, **AS3*** and **AS4***) have smaller differences between TBVs for reduced FD than methods that try to eradicate the black wool gene altogether (**AS2, AS3** and **AS4**). However, within the two selection options, that is, those involving minimising *b* allele frequency, and those involving minimising *bb* expression, the order of the methods changes. When the aim is remove the *b* allele, the option with the least impact on FD is **AS4**, however, when the aim is

to minimise expression of black lambs and maximising gains in FD, the best option is **AS3***. Therefore, optimum methods to minimise expression need less weight on the genotype than methods to eradicate the black wool gene.

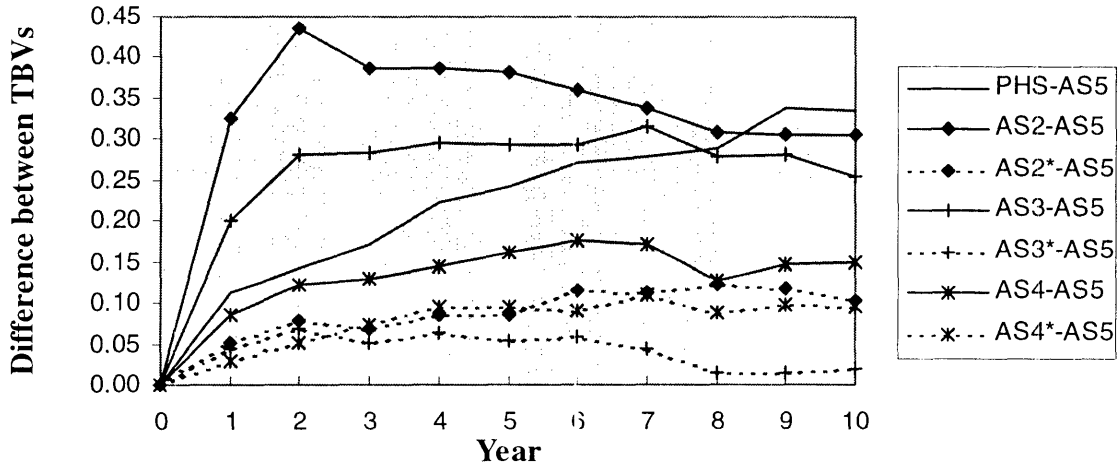

Figure 6.1 Difference in true breeding value (TBV) between combined selection options (**PHS, AS2, AS3, AS4, AS2*,AS3*,AS4***) and selection for FD EBV (AS5) only.

6.3.3 Medium Cost Selection Options

Genotyping males only had similar effects on proportion of animals in each genotype class for the reference selection options **MS1** and **MS1*** (Table 6.7). In general, it took longer to remove the *b* allele in options **MS2**, **MS3** and **MS4** (Table 6.10), and there was the same decrease in heterozygotes apparent with options (**MS2***, **MS3*** and **MS4***) involving reducing the incidence of *bb* animals (Table 6.11). There was a marked increase in incidence of black sheep in the flock with all selection options except **MS2** when compared to equivalent **AS** options. In particular, options **MS4** and **MS4*** were not able to avoid black lamb progeny over the 10 year selection period and **MS4** was not able to reduce the numbers of black lambs any better than **MS4***. This is due to the objective of reducing *bb* animals being less well reached than the objective of reducing the frequency of *b* alleles altogether, when only males are genotyped.

Table 6.10 Results of combined selection on genotype probabilities aimed at minimising frequency of *b* allele (P_{ij}) and estimated breeding value for fibre diameter (EBV_{ij}) with only male animals genotyped

Year	MS2[a]			MS3[b]			MS4[c]		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02
1	29.22	0.08	20.33	33.92	1.77	20.24	44.58	6.17	20.15
2	21.14	0.00	19.69	23.42	0.27	19.61	32.83	3.17	19.51
3	15.14	0.00	19.09	16.82	0.05	19.03	24.87	1.78	18.93
4	9.78	0.00	18.52	11.66	0.00	18.42	19.45	1.20	18.35
5	7.05	0.00	17.93	7.98	0.00	17.84	14.85	0.80	17.75
6	4.57	0.00	17.34	5.15	0.00	17.26	10.51	0.32	17.18
7	2.93	0.00	16.76	3.54	0.00	16.68	7.41	0.15	16.58
8	2.13	0.00	16.19	2.17	0.00	16.10	5.62	0.15	15.98
9	1.47	0.00	15.60	1.42	0.00	15.52	4.18	0.05	15.39
10	1.17	0.00	15.01	1.14	0.00	14.92	3.32	0.04	14.79

[a]weight on $\hat{P}_{ij}$: weight on EBV_{ij} = 2:1
[b]weight on $\hat{P}_{ij}$: weight on EBV_{ij} = 1:1
[c]weight on $\hat{P}_{ij}$: weight on EBV_{ij} = 1:2
[1]*p(b)* mean proportion of *b* allele carriers
[2]*p(bb)* mean proportion of black animals
[3]*mfd* mean FD

Table 6.11 Results of combined selection on genotype probabilities aimed at minimising black lambs ($P_{ij}^{\wedge}$) and estimated breeding value for fibre diameter (EBV_{ij}) with only male animals genotyped

Year	MS2*[a]			MS3*[b]			MS4*[c]		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02
1	43.46	1.83	20.20	51.33	5.03	20.13	58.30	7.91	20.09
2	32.34	0.25	19.59	40.80	2.93	19.51	51.90	7.15	19.45
3	22.43	0.07	19.01	32.07	1.95	18.92	45.41	5.40	18.83
4	15.56	0.09	18.44	24.41	1.07	18.33	38.62	4.35	18.25
5	10.56	0.00	17.88	19.05	0.75	17.73	32.83	3.54	17.65
6	7.25	0.00	17.1	13.55	0.33	17.17	27.69	2.79	17.08
7	4.65	0.00	16.72	9.59	0.10	16.60	23.90	2.37	16.49
8	3.39	0.00	16.15	6.81	0.14	16.05	21.04	2.43	15.91
9	2.43	0.00	15.57	5.02	0.09	15.46	17.74	1.71	15.31
10	1.47	0.00	14.97	3.78	0.07	14.88	15.89	1.66	14.71

[a]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 2:1
[b]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 1:1
[c]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 1:2
[1]*p(b)* mean proportion of *b* allele carriers
[2]*p(bb)* mean proportion of black animals
[3]*mfd* mean FD

When the aim was to reduce frequency of the *b* alleles, greater gains in FD were able to be made using combined selection with only males genotyped (Table 6.10), than when all animals were genotyped (Table 6.8). The cost, however, was the earlier mentioned higher proportions of *Bb* and *bb* animals, than when all animals are genotyped (Table 6.8). Due to the need to select only *BB* sires, there was less gain in FD when only males were genotyped and selection was for reduced expression of *bb* animals (Table 6.9), than when all animals were genotyped.

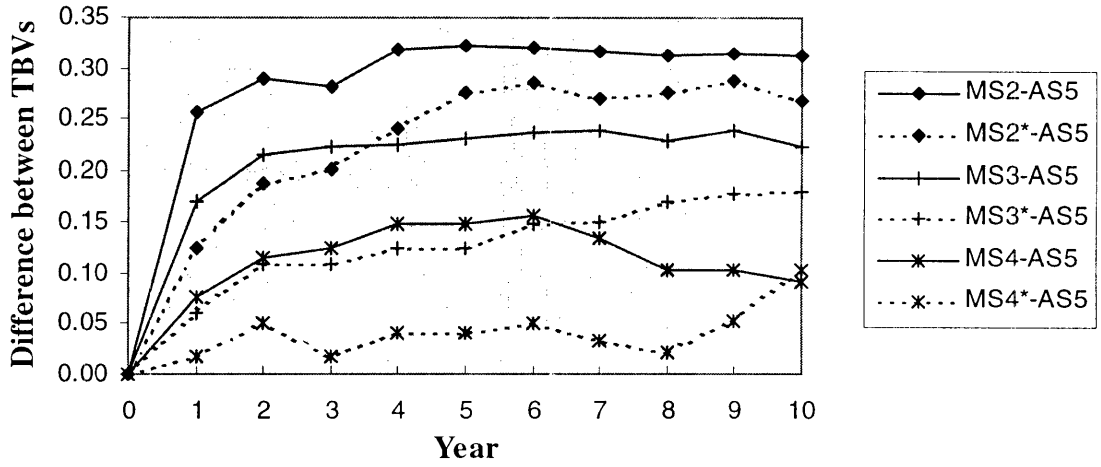

Figure 6.2 Difference in true breeding value (TBV) between combined selection options with only males genotyped (**MS2, MS3, MS4, MS2*, MS3*, MS4***) and selection for FD EBV (**AS5**) only.

The difference between maximum possible gains in FD and gains in FD when only male animals are genotyped are presented in Figure 6.2. This shows that there are not as great differences as when all animals were genotyped (Figure 6.2). There was also not as clear differences between options involving minimising frequency of *b* alleles (**MS2**, **MS3** and **MS4**) and those minimising expression of *bb* animals (**MS2***, **MS3*** and **MS4***).

6.3.4 Low Cost Selection Options

With the aim of reducing the incidence of the *b* allele in the flock while also reducing FD, selection options **PS2**, **PS3** and **PS4** were able to make substantial gains in FD reduction, but were not able to completely remove the *b* allele (Table 6.12). Also, black lambs were present throughout the whole 10 year period of selection and mating. This helped with the reduction of the *b* allele, as having no *bb* alleles would lead to little or no information available to the segregation analysis.

Table 6.12 Results of combined selection on genotype probabilities aimed at minimising frequency of *b* allele (P_{ij}) and estimated breeding value for fibre diameter (EBV_{ij}) with no animals genotyped

Year	**PS2**[a]			**PS3**[b]			**PS4**[c]		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*Mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02
1	35.16	3.95	20.32	41.66	5.75	20.21	49.24	8.99	20.13
2	22.59	1.51	19.77	31.57	3.08	19.59	40.50	6.03	19.50
3	16.88	1.01	19.18	24.70	1.97	19.02	34.90	4.67	18.93
4	12.57	0.58	18.61	18.86	1.21	18.45	29.68	3.57	18.33
5	10.22	0.41	18.02	15.45	1.02	17.87	26.63	3.10	17.74
6	7.13	0.25	17.45	12.04	0.61	17.29	22.29	2.29	17.18
7	5.73	0.26	16.91	10.05	0.50	16.75	19.00	1.95	16.60
8	4.44	0.15	16.33	8.65	0.41	16.15	17.77	1.54	16.00
9	3.03	0.07	15.77	6.20	0.26	15.59	15.40	1.53	15.42
10	2.51	0.10	15.19	5.94	0.23	14.99	14.88	1.45	14.83

[a]weight on $\hat{P}_{ij}$: weight on EBV_{ij} = 2:1
[b]weight on $\hat{P}_{ij}$: weight on EBV_{ij} = 1:1
[c]weight on $\hat{P}_{ij}$: weight on EBV_{ij} = 1:2
[1]*p(b)* mean proportion of *b* allele carriers
[2]*p(bb)* mean proportion of black animals
[3]*mfd* mean FD

In contrast to the results in Table 6.12, with the aim of reducing the frequency of black lambs (*bb* animals) greater gains were able to be made in reduction of FD with selection options **PS2***, **PS3*** and **PS4*** (Table 6.13). This was however, at the cost of a larger proportion of black lambs being born into the flock.

Table 6.13 Results of combined selection on genotype probabilities aimed at minimising black lambs ($P_{ij}^{\wedge}$) and estimated breeding value for fibre diameter (EBV_{ij}) with no animals genotyped

Year	**PS2*[a]**			**PS3*[b]**			**PS4*[c]**		
	p(b)[1]	*p(bb)*[2]	*mfd*[3]	*p(b)*	*p(bb)*	*mfd*	*p(b)*	*p(bb)*	*mfd*
0	64.27	16.37	21.02	64.27	16.37	21.02	64.27	16.37	21.02
1	49.78	4.21	20.18	55.77	5.96	20.11	60.74	7.87	20.09
2	43.13	3.12	19.55	49.39	5.05	19.48	53.99	7.30	19.43
3	36.25	2.49	18.96	44.11	4.50	18.86	50.42	6.54	18.84
4	31.54	2.45	18.38	40.27	4.31	18.28	46.69	5.77	18.25
5	29.11	1.91	17.80	36.11	3.27	17.70	44.24	5.45	17.67
6	26.65	1.68	17.24	33.66	3.39	17.13	41.02	5.53	17.11
7	25.31	1.69	16.65	30.88	2.88	16.56	39.28	4.95	16.54
8	23.04	1.71	16.05	28.72	2.35	15.98	38.31	5.12	15.98
9	20.30	1.41	15.47	27.48	2.23	15.37	35.80	4.39	15.37
10	18.50	1.25	14.87	24.93	2.02	14.77	33.82	3.49	14.78

[a]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 2:1
[b]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 1:1
[c]weight on $P_{ij}^{\wedge}$: weight on EBV_{ij} = 1:2
[1]*p(b)* mean proportion of *b* allele carriers
[2]*p(bb)* mean proportion of black animals
[3]*mfd* mean FD

Results of comparisons between maximum (**AS5**) and realised gains in FD, when no genotyping was available for the black wool gene are shown in Figure 6.3. This shows that the selection options giving the best results for decreased TBV for FD are options **PS4**, **PS4*** and **PS3***. These, however, are the options which result in the greatest proportion of black lambs being born throughout the selection period (Table 6.12 and Table 6.13).

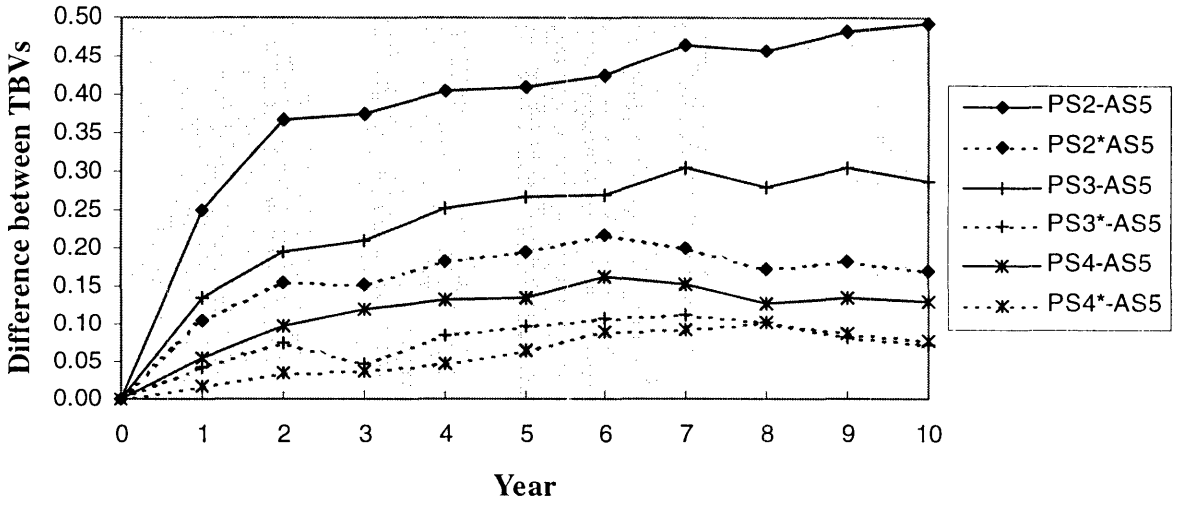

Figure 6.3: Difference in true breeding value (TBV) between combined selection options involving segregation analysis to obtain genotype probabilities (**PS2, PS3, PS4, PS2*, PS3*, PS4***) and selection for FD EBV (**AS5**) only.

6.4 Discussion

The results from this study allow comparisons between different methods for removal of an essentially deleterious recessive allele from a breeding population while continuing to make gains in traits of the breeding objective. This is possible through the use of mate selection - the selection of potential parents and the allocation of mates to optimise progeny genotypes. Comparisons shown in these results are those between more information due to more genotyping (and therefore greater costs) and those due to better methods (segregation analysis and mate allocation). For reference, maximum possible gains in both the reduction of the black gene and decreasing FD were shown in Table 6.6. Any use of the following results relies on comparison of traditional selection methods with maximum achievable gains in the traits of interest.

Comparison of results from traditional selection and individual selection on genotype probabilities and estimated breeding value for FD (with all animals genotyped for the black wool gene) show that traditional selection methods can achieve close to maximum gains in FD over the 10 selection period studied. The difference in achievable FD between these

methods was only 0.34 microns in animals born in year 10. However, traditional selection methods result in black lambs continually being born into the flock, whereas, use of mate selection allows for no black lambs to be born after the first year of selection. Also, in traditional selection strategies there is usually greater loss of selection intensity for FD, as not only are black lambs culled from the flock, their parents and full-sibs are usually removed as well. This strict culling procedure was not implemented in the traditional selection option simulated (**PHS**).

6.4.1 Impact of Genotyping all Animals

Having 100% correct knowledge of black gene genotypes of all animals is a high cost, but results in very little loss of animals due to them being black and therefore little loss in selection intensity for FD. Selection options **AS2**, **AS3**, **AS2*** and **AS3*** all result in no black lambs after year four (and less than 1% after year one). Of these options, the best gains in FD were in from options **AS2*** and **AS3***, both of which allowed heterozygote black gene carriers in the flock. A further selection option which could be investigated involves genotyping all animals and selecting for the removal of the *b* allele. Following complete eradication of the allele, genotyping would no longer be necessary. This reduces the cost of genotyping selection alternatives.

6.4.2 Impact of Genotyping Only Males

With only male animals genotyped (therefore reducing costs) the options involving removal of the black gene all together were more efficient than those allowing heterozygote carriers in the population at reducing numbers of black lambs being born. In these scenarios, **MS4** and **MS4*** produced the best gains in FD. The cost of both of these options, however, was that more black lambs were present in the population than options **MS2**, **MS2***, **MS3** and **MS3***. The cost of having black lambs in the flock must be weighed up against the financial benefit of decreased FD. These were however, less than 2% black lambs after three years of selection for all of these options except **MS4***.

6.4.3 Impact of No Genotyping

Option **PS2*** produced similar gains to options involving genotyping of males only however no genotyping was required. There were still a small percent of black lambs being born in the flock with this selection option, however, gains in FD were comparable. The benefit of this selection option is that there are no genotyping costs. Requirements are simply pedigree and phenotypic performance records for expression of black wool and FD. This option produces less black lambs than natural selection and greater gains in FD for no greater cost.

6.4.4 Simulation Performance

The simulated pedigree and phenotypic information used in this study included information on only one trait from the breeding objective. In practice, stud breeders are selecting animals on a combination of many more traits. In particular selection in this study was for reduced FD, it is well known that there is a positive correlation between fleece weight and FD (Atkins, 1996). Selection for reduced FD will result in decreased fleece weight, unless it is held constant. This was an assumption of the simulated pedigree that will result in less genetic gain in FD than shown in these results. A further assumption was that there is no correlation between FD and black wool. If there was a negative correlation between FD and pigmented fibres the response gained would be slowed down, if there was a positive correlation the response would be sped up. The simulated population represented a stud situation with a closed nucleus, that is no new genes were being brought into the flock. There would be more uncertainty with untested rams being bought into the flock, however, this could be overcome by genotyping all rams bought from outside the stud.

The black gene was simulated in the base population with 35% of white sheep being carriers of the black allele. As mentioned earlier, there are only an average of 6% of white sheep in the industry that carry the allele (Fleet *et al.*, 1995). This is an average and will vary between flocks. The higher proportion of carriers was chosen to illustrate the problems when the black gene is segregating in flocks with higher incidence than average.

The black gene is a simple case of a far larger problem in the wool industry. While black lambs are easily identified, dark or pigmented fibres within the fleece cause similar losses in

wool quality, but are harder to identify. The heritability of such pigmentation is moderate to high (Fleet, 1996) and is thought to be caused by a number of genes. Without detection of these genes, mate selection using segregation analysis would be a useful tool to control expression of pigmentation, although segregation analysis may be difficult if the trait is controlled by many genes. Because of great difficulty in measuring incidence of black fibres, direct DNA tests for this condition would be of immense value.

The mate selection options used in this study included arbitrary weights on the combinations of FD EBV and genotype probability of the progeny. These weights could be determined by a number of different methods. A number of studies have looked at different mate selection strategies to regulate inbreeding while making genetic gain in traits of the breeding objective (Klieve *et al.*, 1994; Wray and Goddard, 1994; Bunter, 1995; Meuwissen, 1997). The genotype probabilities used in this study are similar to inbreeding co-efficients in that the cost of inbreeding is also difficult to weight against the cost of genetic gains for traits of the breeding objective. Inbreeding was not used in this study, but could easily be added to the paired merit functions given in Table 6.4 as an additional restriction. The restriction on expression of the black gene in this study is similar to the restrictions put on inbreeding levels in the studies mentioned above. The arbitrary weights used in this study could be easily adapted to constraining the predicted numbers of black lambs born each year.

6.5 Conclusions

For an observable trait caused by a single recessive gene, such as the gene for black lambs, there is little benefit in genotyping all animals. Similar gains can be made in traits in the breeding objective by use of segregation analysis, or by genotyping only the males, and using mate selection to optimise genetic value of the progeny. In practice this may require greater management demands on the breeder, such as single-sire mating paddocks and selection of individual ewes to be mated to selected rams. However, gains in FD while reducing losses due to black lambs appear substantial. A cost - benefit analysis would have to be done to compare costs of loss of black wool animals and of genotyping with the financial gains possible due to increased response in FD.

Chapter 7

General Discussion

7.1 Introduction

The objective of this study was to examine various aspects of QTL detection and utilisation, with application to Australian Merino sheep breeding. To achieve this aim the previous four studies were completed covering the following areas of QTL and linked marker use in animal breeding: methods to incorporate genetic marker information in the GRM for estimating random QTL effects for individual animals, testing the effect of using selected animals in QTL-marker linkage detection experiments on parameter estimates, incorporating marker information in genotype probability estimates and use of genotype information in mate selection decisions. Important aspects of these studies are discussed in this final chapter including implications of this work to the Australian Wool Industry and general application of genetic marker information within this industry.

Common to the first three experimental chapters of this thesis is the use of genetic marker information to estimate random marker-associated QTL effects. Key to including marker information in the usual mixed model equations for estimating genetic effects is the GRM (Fernando and Grossman, 1989). Construction of the GRM and its inverse was examined in Chapter Three. As shown in this chapter, there is a realised GRM for any one pedigree. This contains only 0's and 1's as elements, as each pair of QTL alleles are either identical or not identical to those of the animal's sire and dam. It was shown that repeating such realisations

of a GRM can be used to simulate true GRM probabilities for a given pedigree, and therefore to test methods of building the GRM. Results of this study showed little differences between methods to build elements of the GRM. Of the methods tested (Fernando and Grossman, 1989; van Arendonk *et al.*, 1994c and Wang *et al.*, 1995) for the given pedigree it is obvious that Wang *et al.* (1995) provide the method of choice to build the GRM and its inverse.

Having established the basis of including random QTL effects in mixed model equations, the same model of Fernando and Grossman (1989) was used in Chapter Four to detect linkage between QTL and genetic markers. The effect of selection in data used for linkage analysis was tested using the model and method of van Arendonk *et al.* (1997) of restricted maximum likelihood to estimate linkage parameters. Within a granddaughter design population, use of selected grandsires has little effect on parameter estimates. However, use of selected sires within grandsire families causes large underestimates of heritability. In general, DNA should be kept from all sires which have progeny test results. This allows the granddaughter design to be used without concern about only having DNA from selected sires and the influence this has on parameter estimates. These results are encouraging for use in sheep breeding, they indicate that QTL mapping can be applied to commercial populations and not only specifically designed populations, provided information is available on the progeny of all sires (both selected and unselected) within grandsire families.

Of further interest to Merino sheep breeders is the ability to estimate QTL genotype probabilities, both with and without the use of genetic marker information. Chapter Five involved the development of a non-iterative genotype probability estimation method. The proposed method did not provide as accurate genotype probability estimates as the iterative method it was tested against (Meuwissen and Goddard, 1997). However, with highly looped pedigrees the iterative method may not produce as accurate estimates as the non-iterative methods and will definitely be slower in computation. Genotype probabilities are not very useful for QTL, since optimising selection for QTL could be in a mixed model framework (polygenic + QTL EBV). Genotype probabilities are useful for single gene traits such as the polled or black gene in Merino sheep. Such a use of genotype probabilities in mate selection decisions was examined in Chapter Six. For major genes, such as those for pigmentation and polled, genotype probabilities from segregation analysis on selected individuals may aid selection decisions for traits of the breeding objective. The use of genetic marker data in

genotype probabilities, however, is not useful for traits, such as those mentioned above, that are highly heritable. Markers will be more useful for traits that can only be measured on one sex (such as reproduction rate), traits that are measured later in life (such as wool production traits) and traits that are difficult to measure (such as resistance to internal and external parasites).

7.2 Detection of QTL-marker linkage in the Wool Industry

Research is underway to detect QTL affecting a number of traits of importance to Australian Merino sheep breeders. These include detection of genes influencing fibre pigmentation (Parsons *et al.*, 1997a), detection of major genes influencing internal parasite resistance (Woolaston *et al.*, 1990), detection of major genes influencing wool protein synthesis (Parsons *et al.*, 1994a, 1994b, 1994c) and detection of genes influencing reproduction characteristics (Montgomery *et al.*, 1993, 1994). It is likely that these genes, or markers closely linked to genes for these traits will be detected in the near future. The question, then becomes how will these genes be used in breeding programs to improve quality and production efficiency of Merino wool production?

Genetic markers can be used to more accurately estimate breeding values for traits of the breeding objective, or alternatively can be used to estimate genotype probabilities for genes at a single locus. The methods available for both of these alternatives have been examined throughout this thesis. The theory and technologies are available for use of molecular marker information in breeding programs following use of linkage and segregation analysis (Chapter 2, Section 2.5). Of further interest is: What traits will marker information be most useful and what breeding structure would be required for detection experiments to detect the presence of segregating QTL in existing breeding populations?

Detection of linkage between genetic markers and quantitative trait loci may be possible within Merino sheep breeding families using a granddaughter design population (Weller *et al.*, 1990). It is even possible to carry out such an analysis of linkage when selected animals are included in the experimental sample (as shown in results from Chapter Four). The use of such a design in the detection of linkage between genetic markers and QTL in the Australian sheep industry may be possible through a progeny testing or sire evaluation scheme. The feasibility

of such an analysis would depend, however, on the total number of animals required for both measurement of phenotypic performance and genotyping. Moody *et al.* (1997) examined the use of the granddaughter design for QTL detection in existing beef cattle populations. QTL of moderate to large effect were able to be detected in populations of Angus and Hereford cattle which had a large number of animals (greater than 100,000). In populations of beef cattle with smaller numbers of animals available (less than 15,000), only QTL of large effect were able to be detected.

The detection of QTL-marker linkage in the wool industry requires genotyping and performance recording information over a number of generations. This type of data structure provides the within family variation required for the detection of linkage. Current industry practice of Merino Sire Evaluation Schemes to compare the performance of progeny from rams from a range of studs at a central location may be the site for linkage analysis studies, because it provides a relatively large number of offspring for individual sires Alternatively, use of parent studs and their associated daughter and multiplier studs in combination with AI and MOET (Brash, 1994) might provide the numbers of animals required. The major problem with the Sire Evaluation Schemes continues to be the small number of sires that are tested each year. It has only been practical to test 12-16 each year. This is a major drawback if QTL-marker linkage detection is carried out. There is, however, the option to combine results over years. This has been done by Cottle *et al.* (1993) with results from 1987-1991. This increased the total number of sires to 92 and gave over 3000 progeny records. This magnitude of information should be more than adequate to test for the presence of QTL given genetic marker data if all 3000 progeny were genotyped (van der Beek *et al.*, 1995; Moody *et al.*, 1997).

In terms of application to nucleus breeding schemes this is a large number of animals to be involved in a trial. However, as results from Chapter Four of this thesis have shown, these animals may be a combination of families from commercial enterprises with only paternal pedigrees and grandsire and sire genotyping required. Van der Beek *et al.* (1995) found that within a three generation family structure, the power of an experiment increased more by doubling the number of offspring per grandsire, than by doubling the number of grandoffspring per grandsire. This may be the way to set up detection experiments in the Merino sheep industry, or to take molecular samples from existing populations, however, this

does rely on adequate numbers of animals available for genotyping and with production records.

7.3 General Conclusions

This thesis combines information on areas of QTL detection and utilisation. It has provided answers to questions relating to the practical application of molecular marker information in animal breeding programs. Specific conclusions resulting from this study are:

- There is little difference in the methods used to build the GRM and its inverse allowing for the most computationally feasible method to be used.
- Most of the variation used in the granddaughter design for linkage analysis studies comes from within family and hence selection of sires (within grandsire families) will have a large impact on the ability to detect linkage between genetic markers and QTL
- By not accounting for selection of sires, within grandsire families, in linkage analysis studies underestimates of parameters such as heritability will occur. Use of information from selected grandsires and granddams, however, has little impact on parameter estimates in linkage analysis studies.
- More research is suggested into the reason why there is no relative change in the proportion of genetic variance due to polygenes and QTL with sire selection in a granddaughter design linkage analysis study. No study has been carried out investigating the effects of selection on both QTL and polygenic variance and the ratio of the two in a detection framework.
- Major gene information in the Merino wool industry may be most useful for traits that are not already included in the breeding objective, such as the recessive gene for black lambs. Use of major gene information will allow the reduction of incidence of black lambs without loosing selection intensity for traits of the breeding objective, such as fleece weights and fibre diameter. While genetic marker information may provide increased accuracy of genotype probabilities, the cost associated with gaining the marker information will not be covered by the gains in accuracy.
- The use of genotype probabilities without marker information can assist breeders in selection decisions for relatively low additional cost.